

2. Data File Variables and Coding Schemes

The rectangular data file included in the *AIDS Public Information Data Set* contains one line of data for each AIDS case reported to CDC. Each line contains 72 columns. The columns contain 49 variables extracted from CDC's national AIDS data set. Please note that some of the variables in the December 1994 data set are located in different columns than in previous versions. The variable *categ* is now in column 4, *central* in column 20, *imm_sev* in column 23, and *MSA* is in column 72. The variable *nir* has been deleted from the data set.

Column	Variable	Description
1	<i>age</i>	Age group at diagnosis of the first AIDS-indicator opportunistic condition
2	<i>sexclass</i>	Sexual classification of patient
3	<i>race</i>	Race of patient
4	<i>categ</i>	Indicates which of the CDC AIDS case revisions the patient meets
5-8	<i>dxdate</i>	Month of diagnosis of first AIDS-indicator opportunistic condition
9-12	<i>repdate</i>	Date when CDC first received information about the case
13	<i>death</i>	Vital status of the patient
14-17	<i>deathqtr</i>	Quarter of death for patients reported dead
18-19	<i>exposure</i>	Mode of exposure to HIV
20	<i>central</i>	Central or outlying country within a metropolitan area
21	<i>multrisk</i>	Indicates if patient had more than one risk of exposure to HIV
22	<i>birth</i>	Country of birth
23	<i>imm_sev</i>	Severe immunosuppression
24	<i>bact</i>	Bacterial infections, multiple or recurrent (including <i>Salmonella</i> septicemia). Applicable in pediatric cases only.
25	<i>burkl</i>	Lymphoma, Burkitt's (or equivalent term)
26	<i>candesop</i>	Candidiasis, esophageal
27	<i>candlung</i>	Candidiasis, bronchi, trachea, or lungs
28	<i>cmv</i>	Cytomegalovirus disease (other than in liver, spleen, or nodes); onset at > 1 month of age
29	<i>cmvret</i>	Cytomegalovirus retinitis (with loss of vision)
30	<i>cocci</i>	Coccidioidomycosis, disseminated or extrapulmonary
31	<i>cryptoco</i>	Cryptococcosis, extrapulmonary
32	<i>cryptosp</i>	Cryptosporidiosis, chronic intestinal
33	<i>dementia</i>	HIV encephalopathy
34	<i>histo</i>	Histoplasmosis, disseminated or extrapulmonary
35	<i>hs</i>	Herpes simplex: chronic ulcer(s) (>1 month duration); or bronchitis, pneumonitis, or esophagitis
36	<i>ibl</i>	Lymphoma, immunoblastic (or equivalent term)
37	<i>iso</i>	Isosporiasis, chronic intestinal (> 1 month duration)
38	<i>ks</i>	Kaposi's sarcoma
39	<i>lip</i>	Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia. Applicable in pediatric cases only.

40	<i>mavium</i>	<i>Mycobacterium avium</i> complex or <i>M. kansasii</i> , disseminated or extrapulmonary
41	<i>myco</i>	<i>Mycobacterium</i> , of other species or unidentified species, disseminated or extrapulmonary
42	<i>pc</i>	<i>Pneumocystis carinii</i> pneumonia
43	<i>plb</i>	Lymphoma, primary in brain
44	<i>pml</i>	Progressive multifocal leukoencephalopathy
45	<i>sals</i>	Salmonella septicemia. Applicable in adult cases only.
46	<i>tb</i>	<i>M. tuberculosis</i> , disseminated or extrapulmonary
47	<i>tp</i>	Toxoplasmosis of brain, onset at >1 month of age
48	<i>wasting</i>	Wasting syndrome due to HIV
49	<i>cervdis</i>	Invasive cervical cancer
50	<i>pulm_tb</i>	Pulmonary tuberculosis
51	<i>rp</i>	Recurrent pneumonia
52	<i>sexbi</i>	Sex with a bisexual man (women only)
53	<i>sexiv</i>	Sex with an injecting drug user
54	<i>sexother</i>	Sex with a person with hemophilia or with a transfusion recipient
55	<i>sexhiv</i>	Sex with a person known to be infected with HIV or to have AIDS, but whose mode of exposure is unknown
56-59	<i>deathrep</i>	Date when death was reported to CDC
60-65	<i>adjwgt</i>	Reporting delay adjustment weight
66-71	<i>dthwgt</i>	Death reporting delay adjustment weight
72	<i>msa</i>	Region of residence at diagnosis of AIDS

Each of these variables is coded alpha-numerically. For example, column 13 contains either “0” or “1.” These numbers represent the variable *death*. The number “0” in this column indicates that CDC has not received a death notification for this case. A value of “1” indicates that CDC has been notified that this patient died. The codes used in the *AIDS Public Information Data Set* are printed below.

Age (column 1)

This variable contains the patient’s age when he or she was first diagnosed with an AIDS-indicator disease.

- 0 = Less than 1 year old
- 1 = 1 to 12 years old
- 2 = 13 to 19 years old
- 3 = 20 to 24 years old
- 4 = 25 to 29 years old
- 5 = 30 to 34 years old
- 6 = 35 to 39 years old
- or age is missing
- 7 = 40 to 44 years old

8 = 45 to 49 years old
9 = 50 to 54 years old
A = 55 to 59 years old
B = 60 to 64 years old
C = 65 years old or older

Sexclass (column 2)

Adult/adolescent males are classified according to their sexual orientation.

1 = Adult/adolescent male who has sex only with other men or sex is missing
2 = Adult/adolescent male who has sex with both men and women
3 = Adult/adolescent heterosexual male or pediatric male
4 = Female (both adult/adolescent and pediatric)

Race (column 3)

1 = White (not Hispanic)
2 = Black (not Hispanic)
3 = Hispanic
4 = Asian/Pacific Islander
5 = American Indian/Alaskan Native
9 = Unknown

Categ (column 4)

This variable reflects changes made over time to the CDC surveillance definition for AIDS. Only cases meeting the current (1993) surveillance definition are included in this data set. *Categ* indicates whether the patient also met the pre-1985, 1985, or 1987 surveillance definition, and whether the diagnosis, if it meets the 1987 or 1993 definition, was definitive or presumptive. Cases that meet more than one of these surveillance definitions are classified into the category listed first. For more information about the 1993 definition, see *Morbidity and Mortality Weekly Report*, December 18, 1992, Recommendations and Reports.

1 = Case meets the pre-1985 surveillance definition
2 = Case meets the 1985 surveillance definition
3 = Case meets the 1987 surveillance definition and was diagnosed definitively
4 = Case meets the 1987 surveillance definition and was diagnosed presumptively
5 = Case meets the 1993 surveillance definition: pulmonary tuberculosis, recurrent pneumonia, and/or cervical cancer (definitive diagnosis)

- 6 = Case meets the 1993 surveillance definition: pulmonary tuberculosis and/or recurrent pneumonia (presumptive diagnosis)
7 = Case meets the 1993 surveillance definition, severe HIV-related immunosuppression

Dxdate (columns 5 through 8)

This variable contains the year and month in which the first AIDS-indicator condition was diagnosed. Columns 5 and 6 contain the year; columns 7 and 8 contain the month. Cases diagnosed before 1982 are coded as “8199.” Cases whose month of diagnosis is unknown are coded as “99” in the month portion of this variable.

Reptime (columns 9 through 12)

This variable contains the year and month in which CDC received the case report. Columns 9 and 10 contain the year; columns 11 and 12 contain the month. Cases reported during 1981 are coded as “8199.”

Death (column 13)

- 0 = CDC has not received a death notification for this case
1 = CDC has been notified that this patient died

Deathqtr (columns 14 through 17)

For patients whose death has been reported to CDC, this variable contains the year and quarter of death. Columns 14 and 15 contain the year; columns 16 and 17 contain the quarter. For example, the value “8803” indicates that the patient died in July, August, or September, 1988.

Exposure (columns 18 and 19)

For surveillance purposes, AIDS cases are counted only once in a hierarchy of exposure categories. Persons with more than one reported mode of exposure to HIV are classified in the exposure category listed first in the hierarchy, except for men with both a history of sexual contact with other men and injecting drug use. They make up a separate exposure category. Persons with multiple reported modes of exposure are indicated in the variable *multirisk*.

“Men who have sex with men” cases include men who report sexual contact with other men (i.e., homosexual contact) and men who report sexual contact with both men and women

(i.e., bisexual contact). “Heterosexual contact” cases are in persons who report specific heterosexual contact with a person with, or at increased risk for, HIV infection (e.g., an injecting drug user).

Adults/adolescents born, or who had sex with someone born, in a country where heterosexual transmission was believed to be the predominant mode of HIV transmission (formerly classified as Pattern-II countries by the World Health Organization) are no longer classified as having heterosexually acquired AIDS. Similar to case reports for other persons who are reported without behavioral or transfusion risks for HIV, these reports are now classified (in the absence of other risk information which would classify them into another exposure category) as “no risk reported or identified” (see *Morbidity and Mortality Weekly Report*, March 11, 1994). Children whose mother was born, or whose mother had sex with someone born, in a Pattern-II country are now classified (in the absence of other risk information which would classify them into another exposure category) as “Mother with/at risk for HIV infection: has HIV infection, risk not specified.”

“Risk not reported or identified” cases are in persons with no reported history of exposure to HIV through any of the routes listed in the hierarchy of exposure categories. Risk not reported or identified cases include persons who are currently under investigation by local health department officials; persons whose exposure history is incomplete because they died, declined to be interviewed, or were lost to follow-up; and persons who were interviewed or for whom other follow-up information was available and no exposure mode was identified. Persons who have an exposure mode identified at the time of follow-up are reclassified into the appropriate exposure category.

Adult/adolescent exposure categories

- 1 = Men who have sex with men
- 2 = Injecting drug use
- 3 = Men who have sex with men and inject drugs
- 4 = Hemophilia/coagulation disorder
- 5 = Heterosexual contact with a person with, or at increases risk for, HIV infection
- 7 = Receipt of blood transfusion, blood components, or tissue
- 8 = Risk not reported or identified

Pediatric exposure categories

- 9 = Hemophilia/coagulation disorder
- 10 = Mother with, or at risk for, HIV infection
- 11 = Receipt of blood transfusion, blood components, or tissue
- 12 = Risk not reported or identified

Central (column 20)

This variable is coded for adult/adolescent patients and designates whether the patient resides in a county which is within the central portion of a metropolitan area, or within an outlying portion.

- 0 = County is not in a metropolitan area (population is less than 50,000)
- 1 = County is in the central portion of a metropolitan area
- 2 = County is in the outlying portion of a metropolitan area
- 9 = Patient's area of residence is unknown, or patient is less than 13 years old

Multrisk (column 21)

Multrisk is coded only for adult/adolescent patients (13 years old or older) and indicates if the patient has risk(s) of exposure to HIV other than the one indicated by *exposure*.

- 0 = Patient's only mode of exposure to HIV is that indicated by *exposure*
- 1 = Patient has additional risk(s) of exposure
- 2 = Patient's mode of exposure is not reported or identified

Birth (column 22)

- 1 = Patient was born in the United States or its dependencies and possessions, or place of birth was not specified
- 2 = Patient was born outside the United States

AIDS-indicator conditions (columns 23 through 51)

Columns 23 through 51 contain information about each of the AIDS-indicator conditions listed on the AIDS confidential case report form. Each of these variables is one character long and is coded as follows:

- 0 = AIDS-indicator condition was not diagnosed
- 1 = AIDS-indicator condition was diagnosed definitively
- 2 = AIDS-indicator condition was diagnosed presumptively

Heterosexual risk information (columns 52 through 55)

These variables (*sexbi*, *sexiv*, *sexother*, and *sexhiv*) contain additional exposure information for patients infected heterosexually. All 4 variables are coded as follows:

- 0 = no
- 1 = yes
- 9 = missing/unknown

The variable *sexbi* is coded only for women (for men, the variable contains a blank). All 4 variables contain “9” (missing/unknown) for patients with hemophilia, regardless of whether the exposure information is in fact unknown. This restriction is necessary in order to comply with the Assurance of Confidentiality on page 5. Of the 1,535 AIDS cases reported through June 1991 among adults/adolescents with hemophilia, less than 3 percent also reported heterosexual contact with a person at increased risk for AIDS or HIV infection.

Deathrep (columns 56 through 59)

For patients whose death has been reported to CDC, this variable contains the year and quarter when CDC received the report. Columns 56 and 57 contain the year; columns 58 and 59 contain the quarter. For example, the value “8803” indicates that the patient’s death was reported to CDC in July, August, or September, 1988.

CDC began collecting this variable in October 1987. Deaths reported to CDC before October 1987 are coded as “8700.”

Adjwgt (columns 60 through 65)

This variable contains an adjustment weight which, when used as a weighting variable in a frequency tabulation, produces tabulations of AIDS cases that are adjusted for delays in case reporting (see page 11 for a discussion of delays in reporting). The weights are based on estimated reporting delay distributions that take into account exposure, geographic, and demographic variations in case reporting. The adjustment weights and the resulting tabulations are not reliable for cases diagnosed during the most recent 6 months. The setup menu contains an adjusted weight option. If you select this option, all subsequent tabulations you request will be weighted accordingly.

Dthwgt (columns 66 through 71)

This variable contains an adjustment weight which, when used as a weighting variable in a frequency tabulation, produces tabulations of AIDS deaths that are adjusted for delays in both case and death reporting (see page 11 for a discussion of delays in reporting). The weights are based on estimated reporting delay distributions that take into account exposure, geographic, and demographic variations in case and death reporting. The adjustment weights and the resulting tabulations are not reliable for deaths occurring during the most recent 6 months. The setup menu contains an adjusted death weight option. If you select this option, all subsequent tabulations you request will be weighted accordingly.

MSA (column 72)

Region of residence at diagnosis of AIDS is identified for adult/adolescent patients residing in MSAs with more than 1 million population, according to the latest available official U.S. Bureau of Census estimates. The MSA variable is coded as:

0 = Not in an MSA (population less than 50,000)

1 = Northeast

Bergen-Passaic, N.J.; Boston, Mass.; Hartford, Conn.; Middlesex, N.J.; Monmouth-Ocean City, N.J.; Nassau-Suffolk, N.J.; New York, N.Y.; Newark, N.J.; or Rochester, N.Y.

2 = Central

Chicago, Ill.; Cincinnati, Ohio; Cleveland, Ohio; Columbus, Ohio; Denver, Colo.; Detroit, Mich.; Indianapolis, Ind.; Kansas City, Mo.; Milwaukee, Wis.; Minneapolis-Saint Paul, Minn.; or Saint Louis, Mo.

3 = West

Anaheim, Calif.; Los Angeles, Calif.; Oakland, Calif.; Oklahoma City, Okla.; Phoenix, Ariz.; Portland, Orig.; Riverside-San Bernardino, Calif.; Sacramento, Calif.; Salt Lake City, Utah; San Diego, Calif.; San Francisco, Calif.; San Jose, Calif.; or Seattle, Wash.

4 = South

Atlanta, Ga.; Charlotte, N.C.; Dallas, Tex.; Fort Lauderdale, Fla.; Fort Worth, Tex.; Houston, Tex.; Memphis, Tenn.; Miami, Fla.; Nashville, Tenn.; New Orleans, La.; Orlando, Fla.; San Antonio, Tex.; San Juan, P.R.; or Tampa, Fla.

5 = Mid-Atlantic

Baltimore, Md.; Norfolk, Va.; Philadelphia, Pa.; Pittsburgh, Pa.; or Washington, D.C.

9 = In an MSA with less than 1 million, but more than 50,000 population

blank = region of residence at diagnosis is unknown, or patient is a pediatric case